Introduction

Semi-Markov processes are a generalization of Markov processes, having a proper method of investigation and a field of application. In the 1950s, some practical problems, especially problems of the mass service, compelled researchers to look for an adequate mathematical description. Attempts to apply Markov models to these problems were sometimes unsatisfactory due to unjustified conclusions about exponential distribution of corresponding time intervals. In order to reduce this limitation P. Lévy [LEV 54] and B. Smith [SMI 55] almost simultaneously proposed the class of stepped processes, called semi-Markov processes. For these processes the Markov property with respect to a fixed non-random time instant is not fulfilled in general cases, but they remain Markov processes with respect to some random times, for example, jump times of their trajectories. Therefore, such a process is not a Markov process, although it inherits some important properties of Markov processes.

Recently the theory of stepped semi-Markov processes has been developed more intensively. An exhaustive bibliography of corresponding works was compiled by Teugels [TEU 76] (see also Koroluk, Brodi, and Turbin [KOR 74], Koroluk and Turbin [KOR 76], Cheong, de Smit, and Togels [CHE 73], Nollau [NOL 80]). The list of works devoted to practical use of this model is very long (see Kovalenko [KOV 65], Korlat, Kuznecov, Novikov, and Turbin [KOR 91], etc.).

We will consider a more general class of random processes with trajectories in a metric space which are continuous from the right and have limits from the left at any point of the half-line. These processes, called semi-Markov processes (in the general sense), have the Markov property with respect to any intrinsic Markov time such as the first exit time from an open set or any finite iteration of such times. The class of semi-Markov processes includes as a sub-class all stepped semi-Markov processes, and also all strong Markov processes, among them continuous ones.

Apparently for the first time general semi-Markov processes [HAR 74] were defined in [HAR 71b] under the name of *processes with independent sojourn times*.

These processes were called continuous semi-Markov processes in [HAR 72] according to the property of their first exit streams. For these processes it is impossible or not convenient to use specific Markov methods of investigation such as semi-group theory, infinitesimal operators, differential parabolic equations or stochastic differential equations. Thus, it is necessary to develop new methods or to modernize traditional methods of investigation, which do not use the simple Markov property. Some methods are borrowed, firstly, from works on stepped semi-Markov processes (see Lévy [LEV 54], Pyke [PYK 61], Smith [SMI 55, SMI 66], etc.), and secondly, from works on Markov processes (see Dynkin [DYN 59, DYN 63], and also [KEM 70, BLU 68, MEY 67], etc.). Some ideas on non-standard description of processes are taken from works of [COX 67]. [HAR 69, HAR 71a] were attempts of such a description. An investigation of some partial types of continuous non-Markovian semi-Markov processes can be found in [CIN 79]. Problems of ergodicity for processes with embedded stepped semi-Markov processes are investigated by Shurenkov [SHU 77].

The simplest element of description of the semi-Markov process is the pair: the first exit time and the corresponding exit position for the process leaving some open set. However, from the classical point of view, such a pair is not simple because it is not convenient to describe it by finite-dimensional distributions of the process ([KOL 36], [GIH 71], etc.). One of the problems we solve in this book is to eliminate this difficulty by means of changing initial elementary data on the process. In our case distributions of the above first exit pairs or joint distributions for a finite number of such pairs becomes the initial data. From engineering point of view it is not difficult to make a device which fixes first exit times and their compositions, and values of the process at these times. It makes it possible to describe the process in a simple, economic and complete way. Such an approach also has some theoretical advantages over the classic approach. In terms of the first exits the continuity problem of the process has a trivial solution. In addition, the necessary and sufficient conditions for the process to converge weakly are simpler than in classical terms. The same can be said for time change problems. However, when accepting this point of view, one must be ready to meet difficulties of another kind. Efforts in this direction would hardly be justified if there did not exist a class of processes such that their description in terms of first exit times would be natural and convenient. This is the class of semi-Markov processes. For them joint distributions of a finite number of the first exit pairs are determined by repeatedly integrated semi-Markov kernels. The semi-Markov process possesses the Markov property with respect to the first exit time from an open set. It is a simple example of such times. The most general class of such times is said to be the class of intrinsic Markov times, which plays an important role in theory of semi-Markov processes.

One of the main features of the class of semi-Markov processes is its closure with respect to time change transformation of a natural type, which generally does not preserve the usual Markov property. One of the main problems in the theory of semi-Markov processes is the problem of representing such a process in the form of a Markov process, transformed by a time change. At present a simple variant of this problem for stepped semi-Markov process (Lévy hypothesis) under some additional assumption is solved (see [YAC 68]). In general, this problem is not yet solved.

General semi-Markov processes are the natural generalization of general Markov processes (like stepped semi-Markov processes generalize stepped Markov processes). Both of these are mathematical models of real phenomena in the area of natural and humanitarian sciences. They serve to organize human experience and to extend the prognostic possibility of human beings as well as any mathematical model. Evidently semi-Markov models cover a wider circle of phenomena than Markov processes. They are more complex due to the absence of some simplifying assumptions, but they are more suitable for applications for natural phenomena than Markov processes. The use of semi-Markov models makes it possible to give a quantitative description for some new features of the process trajectories, which are not accessible using the Markov model.

The characteristic feature of the semi-Markov process is a set of intervals of constancy in its trajectory. For example, such simple operation as truncation transforms Wiener process W into non-Markov continuous semi-Markov process W_a^b:

$$W_a^b(t) = \begin{cases} a, & W(t) \le a, \\ W(t), & a < W(t) < b, \ (t \ge 0) \\ b, & W(t) \ge b, \end{cases}$$

where $a < b$. This process contains intervals of constancy on two fixed levels, but the most typical semi-Markov process possesses intervals of constancy on random spatial positions. In general, such intervals are a consequence of the property of so-called "time run" process with respect to the sequence of states (trace) of the process, which has conditionally independent increments. The famous result of Lévy requires for such a time run process a Poisson field of jumps, which turn into intervals of constancy for an original process itself. This theoretical result finds an unexpected interpretation in chromatography (see [GUT 81]) and in other fields in engineering, where flow of liquid and gas in a porous medium is being used. The continuous semi-Markov model for this movement is reasonable due to the very small inertia of particles of the substance to be filtered. Therefore, the assumption about independence of time intervals, which are taken by non-overlapping parts of the filter while the particle is moving through the filter, is very natural. Moreover, the chromatography curve is a record of the distribution of the first exit time of the particle from the given space interval. Well-known interpretations of intervals of constancy of the particle move trajectory are intervals of delay of particles on the hard phase (convertible adsorption). Such a movement is the experimental fact.

A mathematical theory of reliability is another field of application of continuous semi-Markov processes. For example, the adequate description of an abrasive wear can be given by the inverse process with independent positive increments, i.e. a partial case of a continuous semi-Markov process [VIN 90]. The advantage of such a description appears when analyzing some problems of reliability [HAR 99].

General semi-Markov processes and, in particular, continuous ones can find application in those fields, where the stepped semi-Markov processes are applied, because the former is the limit for a sequence of the latter. Another possible field of application appears when one looks for the adequate model for a Markov-type process, transformed by varying its parameters (see [KRY 77]). The class of semi-Markov processes is more stable relative to such transformations than to Markov ones.

There are many problems of optimal control for random processes, which can be reduced to an optimal choice of a Markov time in order to begin a control action (see [ARK 79, GUB 72, DYN 75, KRY 77, LIP 74, MAI 77, ROB 77, SHI 76]). As a rule, such a Markov time is the first exit time of the process from some set of states. However, the consequence of such a control may be loss of Markovness. An example of such a control is time change depending on a position of the system. Let $(\xi(t))$ be a stationary random process, and $(\psi(t))$ be a time change, i.e. a strictly increasing map $\mathbb{R}_+ \to \mathbb{R}_+$, where $\psi(0) = 0$, $\psi(t) \sim t$ $(t \to \infty)$. Let $(\tilde{\xi}(t))$ be the transformed process. Then

$$\int_0^T \tilde{\xi}(t)\, dt = \int_0^{T'} \xi(t)\psi'(t)\, dt,$$

where $T' = \inf \psi^{-1}[T, \infty) \sim T$ $(T \to \infty)$. The average in time of the transformed process can be more than that of the original process, if $\psi'(t)$ is chosen in a suitable form. For example, it would be the case, if $\psi'(t) = f(\xi(t))$, where f is an increasing function. In this case ψ is uniquely determined if the function ξ is determined. The similar effect of enlargement of the average meaning is possible, when under fixed ξ the time change is some non-degenerated random process. For example, let $\psi'(t)$ be a Poisson process with the intensity $f(\xi(t))$ (Cox process [COX 67]). Then

$$\begin{aligned} E \int_0^T \tilde{\xi}(t)\, dt &= E \int_0^{T'} \xi(t)\, d\psi(t) \\ &= E \int_0^{T'} \xi(t) f\big(\xi(t)\big)\, dt \sim \int_0^T \xi(t) f\big(\xi(t)\big)\, dt \quad (T \longrightarrow \infty). \end{aligned}$$

If the first form determines the random time change, then the second form of time change is twice random. The main distinction of these two kinds of time changes is in their relation to Markovness. The former preserves the Markov property and the latter does not. The process of the second type appears to be semi-Markovian, and the class of such processes is invariant with respect to such a time change. An answer on the practical question about preference of the time change form depends on engineering possibilities.

Let us pay attention to another aspect of the problem. In [HAR 90] the possibility of checking the hypothesis on Markovness of a one-dimensional continuous process is discussed. It was noted that to reveal Markovness of a process with an uncountable set of states is not possible in principle, when probability of the process to be at when any partial one-dimensional distribution of the process is continuous. In addition, it is impossible to check this hypothesis for discrete time and continuous set of states, and also that for continuous time and multi-dimensional $(d \geq 2)$ space of states. It follows from the fact that it is impossible to have a reasonable statistical estimate for conditional probability $P(A|\xi(t) = x)$ for a random process ξ, if $P(\xi(t) = x) = 0$ for any $t \in \mathbb{R}$. The last condition can possibly be true for the one-dimensional process too. However, in this case, in order to check Markovness we can use the infinite sequence of random times, when the trajectory of the process crosses the level x. We can organize a sequence of Markov times containing iterated first exit times from interval (∞, x). For a strictly Markov process any such sequence of marked point processes determines a stepped semi-Markov process. The first test must be checking this property of the point process, although it does not guarantee Markovness of the original process. Therefore, if the first test is positive, the second test must check Markovness with the help of distributions of the first exit times from small neighborhoods of an initial point. For a proper Markov process this distribution has a special limit property while diameters of the neighborhoods tend to zero. Later on we consider two corresponding criteria of Markovness for semi-Markov processes.

It seems to be convincing enough that continuous semi-Markov processes are worth investigating. Their practical usefulness is connected with numerical calculations. Here we will not cover computer problems. The main theme of this book is devoted to theoretical aspects of continuous semi-Markov models. It happens that there are a lot of such aspects.

The book is divided into chapters, sections, subsections, and paragraphs. In every chapter there is its own enumeration of divisions, and as well that of theorems, lemmas, and propositions. All these items, except subdivisions and paragraphs, have double numbers. For example, in 3.14 the number 3 means chapter, and 14 the number of this object in this chapter. Subdivisions have triple numbers. For example, 5.2.3 means the follows: 5 is the chapter number, 2 is that of the section, 3 is the number of subsection in the section. Paragraphs have single numbers. A reference on a paragraph inside a given chapter is given as item 9, where 9 is its number. A reference to a paragraph from another chapter is given as item 2.4, where 2 is the chapter number, and 4 is the paragraph number within this chapter.

It is necessary to include in Chapter 1 some information about stepped semi-Markov processes. It will be assumed that the reader is familiar with this concept. The main focus will be on constructing a measure of the stepped semi-Markov process with the help of the method, which will be further generalized for the general semi-Markov processes. In addition, we give without proof the main result of the theory of stepped semi-Markov processes, namely the ergodic theorem and the corresponding formula of their stationary distribution.

In Chapter 2 a method of investigation of a process with the help of embedded point processes (streams) of the first exits relative to a sequence of subsets of the given metric space of states is discussed. Deducing sequences of subsets, the main tool for analyzing properties of random processes connected with the first exit times, are defined. The structure of the set of regenerative times for a measurable family of probability measures is investigated. Such a name is assigned to a set of Markov times for which the Markov property for the given family of measures is fulfilled for.

In Chapter 3 general semi-Markov processes are defined and investigated. Such a process is characterized by a measurable family of probability measures with the set of regenerative times which includes any first exit time from an open set. Semi-Markov transition functions, transition generating functions, and lambda-characteristic operators are considered. The conditions necessary for a semi-Markov process to be Markov are analyzed.

In Chapter 4 the semi-Markov process is constructed on the base of *a priori* given family of semi-Markov transition functions. It is considered as a special case of more general problem how to construct a process from a consistent system of marked point processes, namely, that of streams which can be interpreted as the streams of the first exit pairs (time, position) from open sets for some process.

In Chapter 5 semi-Markov diffusion processes in a finite-dimensional space are considered. They enable the description in terms of partial differential elliptical equations.

In Chapter 6 properties of trajectories of semi-Markov processes are investigated. For any trajectory, a class of trajectories is defined in such a way that a trajectory from this class differs from the origin one only by some time change. This class can be interpreted as a sequence of states which the system goes through without taking into account the time spent on being in these states. We call this the trace of the trajectory. The individual trajectory from this class is distinguished by time run along the trace. The measure of the semi-Markov process induces a distribution on the set of all traces. This distribution possesses the Markov property of a special kind. In this aspect the projection is related to the Markov process. The special property of the semi-Markov process is exposed by its conditional distribution of time run along the trace. The character of this distribution is reflected in properties of trajectories of the semi-Markov process.

In Chapter 7 we consider conditions for the sequence of stepped semi-Markov processes to converge weak to a general semi-Markov process. Results related to semi-Markov processes are obtained as a consequence of general theorems about weak compactness and weak convergence of probability measures in the space $\mathcal{D}$. These theorems in terms of the first exits times have some comparative advantages with the similar results in terms of the usual finite-dimensional distributions.

In Chapter 8 a problem of representation of the semi-Markov process in the form of a Markov process, transformed by a random time change, is solved. We demonstrate two methods of solving this problem with construction of the corresponding Markov process and time change. The first solution is formulated in terms of the so called lambda-characteristic operator of this process. The corresponding Markov process is represented by its infinitesimal operator. The second solution is obtained in terms of Lévy decomposition of the conditional process with independent positive increments (process of time run). The corresponding Markov process is determined by the distribution of its random trace, and that of the conditional distribution by its time run along the trace. These distributions are used to derive formulae of stationary distributions of corresponding semi-Markov processes.

In Chapter 9 we consider some applications of continuous semi-Markov processes. It seems natural that they can serve as an adequate model for carrying substance through a porous medium. This follows from both some premises about independence, and experimental data of such a phenomenon as chromatography. The latter gives the purest example of a continuous semi-Markov process. We develop semi-Markov models for both the liquid and gas chromatography. The similar application has a geological origin. We consider a continuous semi-Markov model of accumulation of a substance which consists of particles moving up to some non-Markov time and remaining in the stop position forever.

According to accepted convention we use denotations like $P(A, B)$ instead of $P(A \cap B)$. We also will omit braces in expression like $A = \{X_n \in C\}$ as an argument of a measure or expectation: $P(A) = P(X_n \in C)$, $E(f; A) = E(f; X_n \in C)$, and so on.

The author would like to express their deep appreciation for the help given by Yu. V. Linnik, who supported the first works of the author on continuous semi-Markov processes. The author expresses sincere gratitude to all participants of the seminar on probability theory and mathematical statistics at St Petersburg Department of Steklov Mathematical Institute of the Russian Academy of Sciences, and especially to I. A. Ibragimov for their assistance with this work.